1

Double Seasonal Recurrent Neural Networks for Forecasting Short Term Electricity Load Demand in Indonesia

Sony Sunaryo, Suhartono and Alfonsus J. Endharta
Department of Statistics, Institut Teknologi Sepuluh Nopember
Indonesia

1. Introduction

PT. PLN (*Perusahaan Listrik Negara*) is Government Corporation that supplies electricity needs in Indonesia. This electricity needs depend on the electronic tool used by public society, so that PLN must fit public electricity demands from time to time. PLN works by predicting electricity power which is consumed by customers hourly. The prediction made is based on prior electricity power use.

The prediction of amount of electricity power use is done to optimize electricity power used by customers, so that there will not be any electricity extinction. There are some methods that could be used for forecasting of amount of electricity power use, such as double seasonal ARIMA model and Neural Network (NN) method. Some researches that are related to short-term electricity power forecasting can be seen in Chen, Wang and Huang (1995), Kiartzis, Bakirtzis and Petridis (1995), Chong and Zak (1996), Tamimi and Egbert (2000), Husen (2001), Kalaitzakis, Stavrakakis and Anagnostakis (2002), Taylor (2003), Topalli and Erkmen (2003), Taylor, Menezes and McSharry (2006), and Ristiana (2008). Neural network methods used in those researches are Feed Forward Neural Network, which is known as AR-NN model. This model cannot get and represent moving average order effect in time series. Some prior researches in many countries in the world including in Indonesia showed that ARIMA model for the electricity consumption data tends to have MA order (see Taylor, Menezes and McSharry (2006) and Ristiana (2008)).

The aim of this research is to study further about other NN type, i.e. Elman-Recurrent Neural Network (RNN) which can explain both AR and MA order effects simultaneously for forecasting double seasonal time series, and compare the forecast accuracy with double seasonal ARIMA model. As a case study, we use data of hourly electricity load demand in Mengare, Gresik, Indonesia. The results show that the best ARIMA model for forecasting these data is ARIMA ([1,2,3,4,6,7,9,10,14,21,33],1,8)(0,1,1)24(1,1,0)168. This model is a class of double seasonal ARIMA, i.e. daily and weekly seasonal with 24 and 168 length of periods respectively. Additionally, there are 14 innovational outliers detected from this ARIMA model.

In this study, we apply 4 different architectures of RNN particularly for the inputs, i.e. the input units are similar to ARIMA model predictors, similar to ARIMA predictors plus 14 dummy outliers, the 24 multiplied lagged of the data, and the combination of 1 lagged and the 24 multiplied lagged plus minus 1. The results show that the best network is the last ones, i.e., Elman-RNN(22,3,1). The comparison of forecast accuracy shows that Elman-RNN yields less MAPE than ARIMA model. Thus, Elman-RNN(22,3,1) gives more accurate forecast values than ARIMA model for forecasting hourly electricity load demands in Mengare, Gresik, Indonesia.

The rest of this paper is organized as follows. Section 2 briefly introduces the forecasting methods, particularly ARIMA and NN methods. Section 3 illustrates the data and the proposed methodology. Section 4 evaluates the model's performance in forecasting double seasonal data and compares the forecasting accuracy between the RNN and ARIMA models. The last section gives the conclusion and future work.

2. Forecasting methods

There are many quantitative forecasting methods based on time series approach. In this section, we will briefly explain some methods used in this research, i.e. ARIMA model and Neural Network.

2.1 ARIMA model

One of the popular time series models and mostly used is ARIMA model. This model contains three parts, namely autoregressive (AR), moving average (MA), and mix of ARMA models (Wei, 2006). Bassically, this model model shows that there is a relationship between a value in the present (Z_t) and values in the past (Z_{t-k}), added by random value. ARIMA (p,d,q) model is a mixture of AR(p) and MA(q), with a non-stationery data pattern and d differencing order. The mathematics form of ARIMA(p,d,q) is

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)\, a_t \tag{1}$$

where p is AR model order, q is MA model order, d is differencing order, and

$$\phi_p(B) = (1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\,,$$

$$\theta_q(B) = (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\,.$$

Generalization of ARIMA model for a seasonal pattern data, which is written as ARIMA (p,d,q)(P,D,Q)^S, is (Wei, 2006)

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)\, a_t \tag{2}$$

where s is seasonal period, and

$$\Phi_P(B^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps})\,,$$

$$\Theta_Q(B^s) = (1-\Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs})\,.$$

Short-term (half-hourly or hourly) electricity consumption data frequently follows a double seasonal pattern, including daily and weekly seasonal. ARIMA model with multiplicative double seasonal pattern as a generalization of seasonal ARIMA model, written as ARIMA(p,d,q)(P_1,D_1,Q_1)^{S1}(P_2,D_2,Q_2)^{S2}, has a mathematical form as

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Phi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}Z_t = \theta_q(B)\Theta_{Q_1}(B^{s_1})\Theta_{Q_2}(B^{s_2})a_t \tag{3}$$

where s_1 and s_2 are periods of difference seasonal.
One of the methods that can be used to estimate the parameters of ARIMA model is Maximum Likelihood Estimation (MLE) method. The assumption needed in MLE method is that error a_t distributes normally (Box, Jenkins and Reinsel, 1994; Wei, 2006). Therefore, the cumulative distribution function is

$$f(a_t \mid \sigma_a^2) = (2\pi\sigma_a^2)^{-1/2} \exp\left(-\frac{a_t^2}{2\sigma_a^2}\right) \tag{4}$$

Because error is independent, the jointly distribution from $a_1, a_2, \ldots, a_n$ is

$$f(a_1, a_2, \ldots, a_n \mid \sigma_a^2) = (2\pi\sigma_a^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_a{}^2}\sum_{t=1}^{n} a_t^2\right). \tag{5}$$

Error a_t can be stated as a function of Z_t, and parameters ϕ, θ, σ_a^2 and also the prior error. Generally a_t is written as

$$a_t = Z_t - \phi_1 Z_{t-1} - \ldots - \phi_p Z_{t-p} + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}\ . \tag{6}$$

The likelihood function for parameters of ARIMA model when the observations are known is

$$L(\phi, \theta, \sigma_a^2 \mid Z) = (2\pi\sigma_a^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_a{}^2} S(\phi, \theta)\right) \tag{7}$$

where

$$S(\phi, \theta) = \sum_{t=1}^{n} (Z_t - \phi_1 Z_{t-1} - \ldots - \phi_p Z_{t-p} + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q})^2\ . \tag{8}$$

Then, the log-likelihood function is

$$l(\phi, \theta, \sigma_a^2 \mid Z) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_a^2) - \frac{1}{2\sigma_a{}^2} S(\phi, \theta)\ . \tag{9}$$

The maximum of the log-likelihood function is computed by finding the first-order derivative of Equation (9) to each parameter and equaling to zero, i.e.

$$\frac{\partial l(\phi, \theta, \sigma_a^2 \mid Z)}{\partial \phi} = 0\ ;\ \frac{\partial l(\phi, \theta, \sigma_a^2 \mid Z)}{\partial \theta} = 0\ ;\ \frac{\partial l(\phi, \theta, \sigma_a^2 \mid Z)}{\partial \sigma_a^2} = 0\ .$$

An information matrix which is notated as $I(\phi, \theta)$ is used to calculate the standard error of estimated parameter by MLE method (Box, Jenkins and Reinsel, 1994). This matrix is obtained by calculating the second-order derivative to each parameter $(\beta = (\phi, \theta))$, which is notated as I_{ij} where

$$I_{ij} = \frac{\partial^2 l(\beta, \sigma_a^2 \mid Z)}{\partial \beta_i \partial \beta_j}, \tag{10}$$

and

$$I(\beta) = -E(I_{ij}). \tag{11}$$

The variance of parameter is notated as $V(\hat{\beta})$ and the standard error is $SE(\hat{\beta})$.

$$V(\hat{\beta}) = [I(\hat{\beta})]^{-1} \tag{12}$$

and

$$SE(\hat{\beta}) = [V(\hat{\beta})]^{1/2}. \tag{13}$$

2.2 Neural Network

In general Neural Network (NN) has some components, i.e. neuron, layer, activation function, and weight. NN modeling could be seen as the network form which is including the amount of neurons in the input layer, hidden layer, and output layer, and also the activation functions. Feed-Forward Neural Network (FFNN) is the mostly used NN model for time series data forecasting (Trapletti, 2000; Suhartono, 2007). FFNN model in statistics modeling for time series forecasting can be considered as a non-linear autoregressive (AR) model. This model has a limitation, which can only represent AR effects in time series data.

One of the NN forms which is more flexible than FFNN is Recurrent Neural Network (RNN). In this model the network output is set to be the input to get the next output (Beale and Finlay, 1992). RNN model is also called Autoregressive Moving Average-Neural Network (ARMA-NN), because the inputs are not only some lags of response or target, but also lags of the difference between the target prediction and the actual value, which is known as the error lags (Trapletti, 2000). Generally, the architecture of RNN model is same with ARMA(p,q) model. The difference is RNN model employs non-linear function to process the inputs to outputs, whereas ARMA(p,q) model uses linear function. Hence, RNN model can be said as the non-linear Autoregressive Moving Average model.

There are many activation functions that could be used in RNN. In this research, tangent sigmoid function and linear function are used in hidden layer and output layer respectively. The mathematics form of tangent sigmoid function is

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}, \tag{14}$$

and linear function is $f(x) = x$. The architecture of Elman-RNN, for example ARMA(2,1)-NN and 4 neuron units in hidden layer, is shown in Fig. 1.

In general, Elman-RNN(2,4,1) or ARMA(2,1)-NN is a nonlinear model. This network has 3 inputs, such as Y_{t-1}, Y_{t-2} and residual e_{t-1}, four neuron units in the hidden layer with activation function $\Psi(\bullet)$ and one neuron in the output layer with linear function. The main

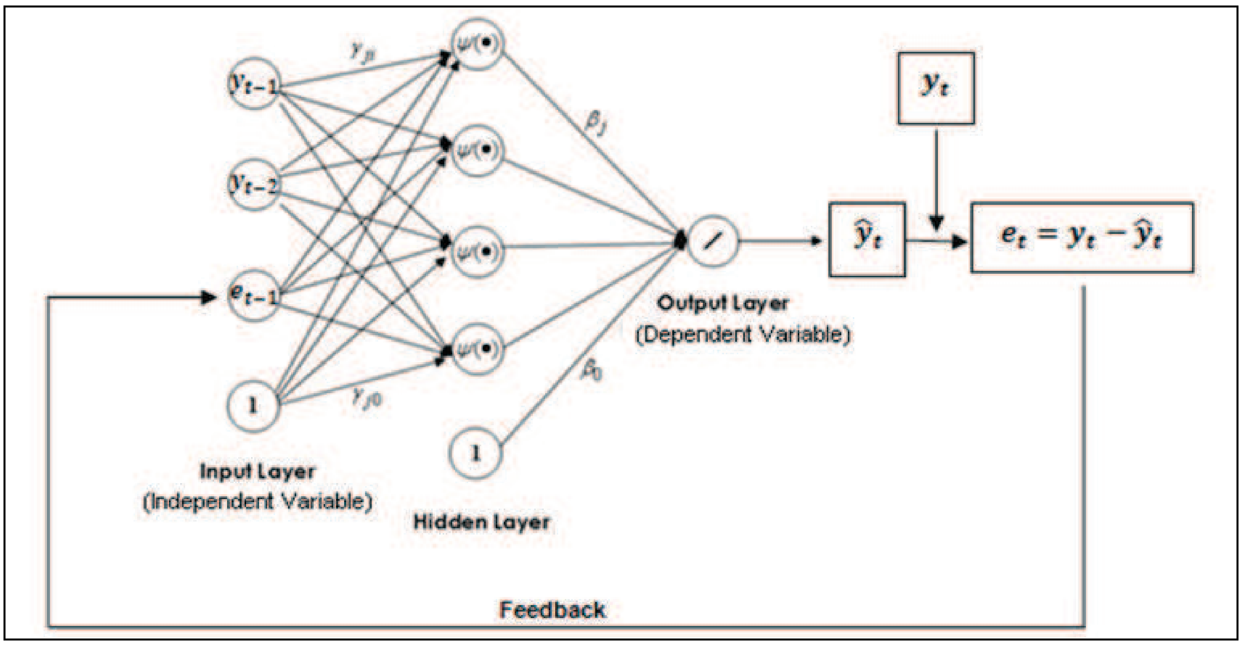


Fig. 1. The architecture of Elman-RNN(2,4,1) or ARMA(2,1)-NN

difference between Elman-RNN and other NN types is the presence of feedback process, i.e. a process representing the output as the next input. Therefore, the advantage of using Elman-RNN is the fits or predictions are usually more accurate, especially for data that consist of moving average order.

The weight and the bias in the Elman-RNN model are estimated by backpropagation algorithm. The general RNN with one hidden layer, q input units and p units in the hidden layer is

$$Y = f^{o}\left[\beta_0 + \sum_{j=1}^{p}\left(\beta_j f^{h}\left(\gamma_{j0} + \sum_{i=1}^{q}\gamma_{ji}X_i\right)\right)\right] \tag{15}$$

where β_j is the weight of the j-th unit in the hidden layer, γ_{ji} is the weight from i-th input to j-th unit in the hidden layer, $f^{h}(x)$ is the activation function in the hidden layer, and $f^{o}(x)$ is the function in the output layer. Chong and Zak (1996) explain that the weight and bias can be estimated by minimizing the value E in the following equation

$$E = \frac{1}{2}\sum_{k=1}^{n}[Y_{(k)} - \hat{Y}_{(k)}]^2 . \tag{16}$$

Minimization of Equation (16) is done by using Gradient Descent method with momentum. Gradient Descent method with momentum m, $0<m<1$, is formulated as

$$w^{(t+1)} = w^{(t)} - \left(m \cdot dw^{(t)} + (1-m)\eta\frac{\partial E}{\partial w}\right) \tag{17}$$

where dw is the change of the weight or bias, η is the learning rate which is defined, $0<\eta<1$. To solve the equation, we do the partial derivative of E to each weight and bias w with chain rules. The partial derivative of E to the weight β_j is

$$\frac{\partial E}{\partial \beta_j} = -\sum_{k=1}^{n} [Y_{(k)} - \hat{Y}_{(k)}] f^{o'}\left(\beta_0 + \sum_{l=1}^{p} \beta_l V_{l(k)} \right) V_{j(k)} \ . \tag{18}$$

Equation (18) is simplified into

$$\frac{\partial E}{\partial \beta_j} = -\sum_{k=1}^{n} \delta^o{}_{(k)} V_{j(k)} \tag{19}$$

where

$$\delta^o{}_{(k)} = [Y_{(k)} - \hat{Y}_{(k)}] f^{o}\left(\beta_0 + \sum_{l=1}^{p} \beta_l V_{l(k)} \right).$$

By using the same way, the partial derivatives of E to β_0 , γ_{li} , and γ_{l0} are done, so that

$$\frac{\partial E}{\partial \beta_0} = -\sum_{k=1}^{n} \delta^o{}_{(k)} \ , \tag{20}$$

$$\frac{\partial E}{\partial \gamma_{ji}} = -\sum_{k=1}^{n} \left[Y_{(k)} - \hat{Y}_{(k)} \right] f^{o'}\left(\beta_0 + \sum_{l=1}^{p} \beta_l V_{l(k)} \right) \times \ \beta_j f^{h'}\left(\gamma_{l0} + \sum_{i=1}^{q} \gamma_{li} X_{i(k)} \right) X_{l(k)} \tag{21}$$

or

$$\frac{\partial E}{\partial \gamma_{ji}} = -\sum_{k=1}^{n} \delta^h{}_{(k)} X_{i(k)} \ , \tag{22}$$

and

$$\frac{\partial E}{\partial \gamma_{j0}} = -\sum_{k=1}^{n} \delta^h{}_{(k)} \ \ , \tag{23}$$

where

$$\delta^h{}_{(k)} = \delta^o{}_{(k)} \beta_j f^{h'} (\gamma_{l0} + \textstyle\sum_{i=1}^{q} \gamma_{li} X_{i(k)}) \ \ . \tag{24}$$

These derivatives process shows that the weight and the bias can be estimated by using Gradient Descent method with momentum. The the weight and the bias updating in the output layer are

$$\beta_j^{(s+1)} = \beta_j^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^o{}_{(k)} V_{j(k)} \right) \tag{25}$$

and

$$\beta_0{}^{(s+1)} = \beta_0{}^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^o{}_{(k)} \right). \tag{26}$$

The weight and the bias updating in the hidden layer are

$$\gamma_{ji}^{(s+1)} = \gamma_{ji}^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^{h}{}_{(k)} X_{i(k)} \right) \tag{27}$$

and

$$\gamma_{j0}^{(s+1)} = \gamma_{j0}^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^{h}{}_{(k)} \right). \tag{28}$$

In Equation (25) to (28), dw is the change of the related weight or bias, m is the momentum, and η is the learning rate.

3. Data and methodology

This research uses an electricity consumption data from Electrics Goverment Company (PLN) in Gresik region as a case study. The data is hourly electricity consumption data in Mengare Gresik, which is recorded from 1 August to 23 September 2007. Then, data are divided into two parts, namely in-sample for observations in period of 1 August to 15 September 2007 and out-sample dataset for 16-23 September 2007. Fig. 2 shows the time series plot of the data.

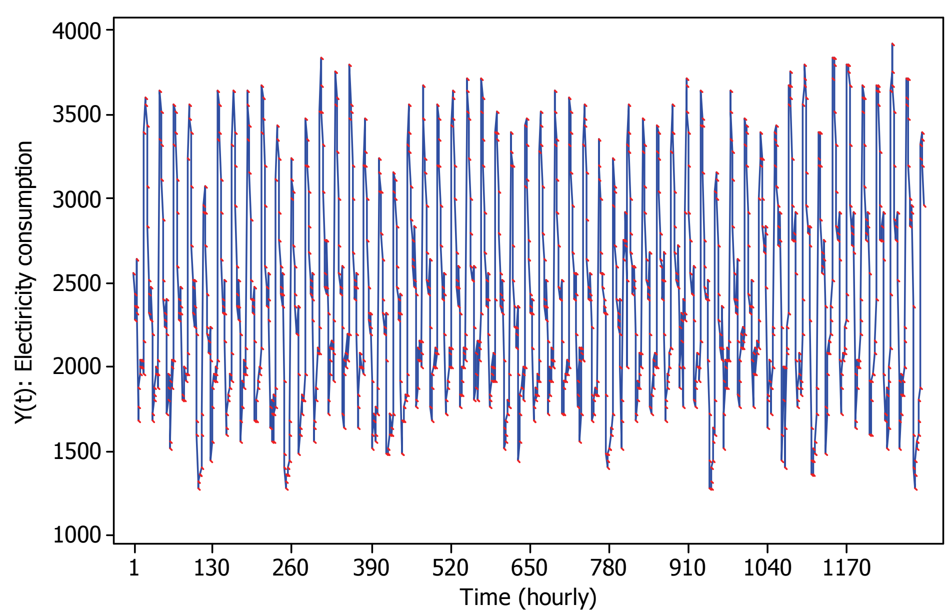


Fig. 2. Time series plot of hourly electricity consumption in Mengare Gresik, Indonesia

The methodology for analysing the data consists of the following steps:

i. Modeling of double seasonal ARIMA by using Box-Jenkins procedure.
ii. Modeling of Elman-RNN with four types of input, i.e.
   a. The inputs are based on the order of the best double seasonal ARIMA model at the first step.
   b. The inputs are based on on the order of the best double seasonal ARIMA model at the first step and dummy variables from outliers detection.
   c. The inputs are the multiplication of 24 lag up to lag 480.
   d. The inputs are lag 1 and multiplication of 24 lag ± 1.

iii. Forecast the out-sample dataset by using both Elman-RNN and double seasonal ARIMA model.
iv. Compare the forecast accuracy between Elman-RNN and double seasonal ARIMA model to find the best forecasting model.

4. Results

A descriptive data analysis shows that the highest electricity consumption is at 19.00 pm about 3537 kW, and the lowest is at 07.00 am about 1665,2 kW. This consumption explains that at 07.00 am most of customers turn the lamps off, get ready for work, and leave for the office. In Indonesia, customer work hours usually begins at 09.00 am and end at 17.00 pm. Thus, the household electricity consumption at that time period is less or beyond of the average of overall electricity consumption. At 18.00 pm, customers turn the night lamps on and at 19.00 pm most of customers have been back from work, and do many kinds of activities at house, that use a large amount of electricity such as electronics use.
Summary of descriptive statistics of the daily electricity consumption can be seen in Table 1. This table illustrates that on Tuesday the electricity consumption is the largest, about 2469.6 kW, and the lowest electricity consumption is on Sunday, about 2204.8 kW. The electricity consumption averages on Saturday and Sunday are beyond the overall average because those days are week-end days, so that customers tend to spend their week-end days with their family outside the house.

Day	Number of observations	Mean	Standard Deviation
Monday	168	2439.0	624.1
Tuesday	168	2469.5	608.2
Wednesday	192	2453.3	584.8
Thursday	192	2447.9	603.9
Friday	192	2427.3	645.1
Saturday	192	2362.7	632.4
Sunday	192	2204.8	660.3

Table 1. Descriptive Statistics of the Hourly Electricity Consumption in Every Day

4.1 Result of double seasonal ARIMA model

The process for building ARIMA model is based on Box-Jenkins procedure (Box, Jenkins and Reinsel, 1994), starting with identification of the model order from the stationer data. Fig. 2 shows that the data are non-stationer, especially in the daily and weekly periods.
Fig. 3 shows the ACF and PACF plots of the real data, and indicate that the data are non-stationer based on the slowly dying down weekly seasonal lags in ACF plot. Hence, daily seasonal differencing (24 lags) should be applied. After daily seasonal differencing, ACF and PACF plots for these differencing data are shown in Fig. 4. ACF plot shows that ACF at regular lags dies down very slowly, and indicates that regular order differencing is needed. Then, daily seasonal and regular order differencing data have ACF and PACF plots in Fig. 5. The ACF plot in this figure shows that lags 168 and 336 are significant and tend to die down very slowly. Therefore, it is necessary to apply weekly seasonal order differencing (168 lags).

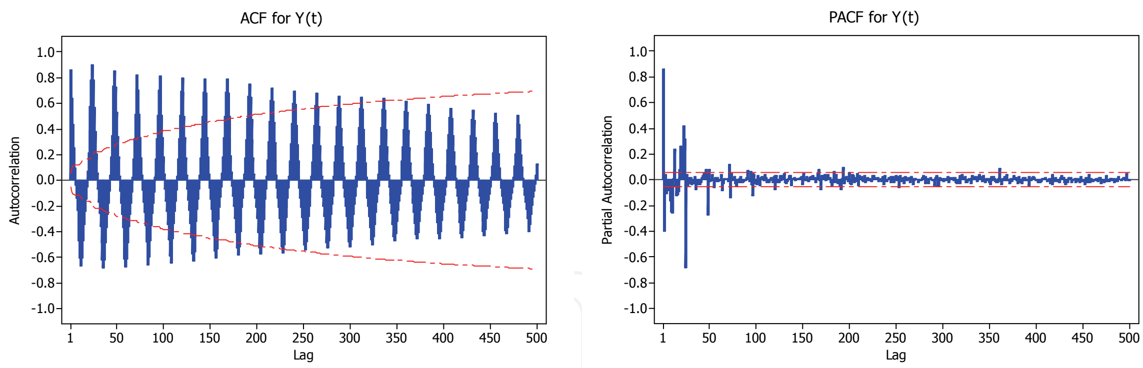


Fig. 3. ACF and PACF for original hourly electricity consumption data

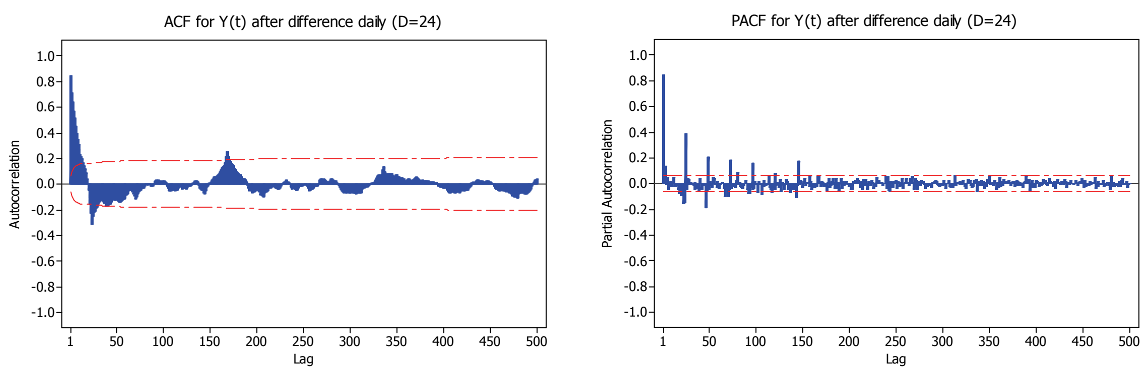


Fig. 4. ACF and PACF for data after differencing daily seasonal (D=24)

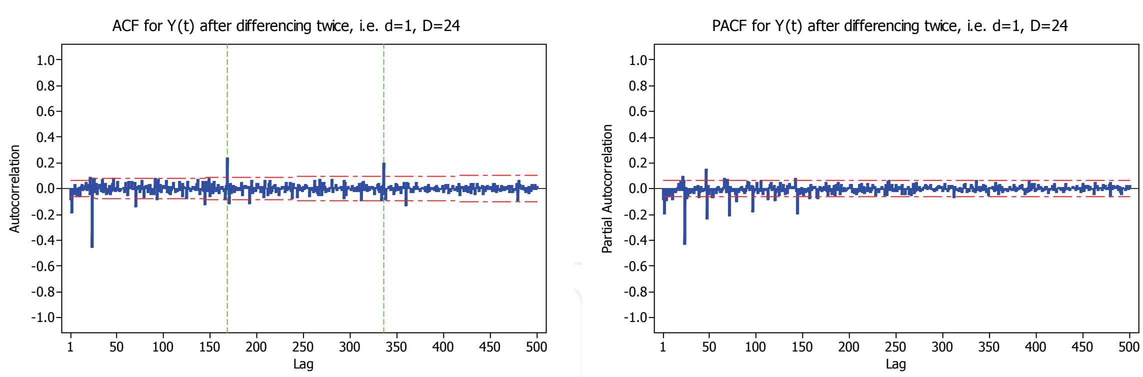


Fig. 5. ACF and PACF for data after differencing twice, i.e. d=1 and D=24

Figure 6 shows that the ACF and PACF plots of stationer data, which are the data that has been differenced by lag 1, 24, and 168. Based on these ACF and PACF plots, there are two the tentative double seasonal ARIMA models that could be proposed, i.e. ARIMA ([1,2,3,4,6,7,9,10,14,21,33],1,[8])$(0,1,1)^{24}(1,1,0)^{168}$ and ([12],1,[1,2,3,4,6,7])$(0,1,1)^{24}(1,1,0)^{168}$. Then, the results of parameters significance test and diagnostic check for both models show that the residuals are white noise. Moreover, the results of Normality test of the residual with Kolmogorov-Smirnov test show that the residuals for both models do not satisfy normal distribution. It due to some outliers in the data and the complete results of outliers' detection could be seen in Endharta (2009).

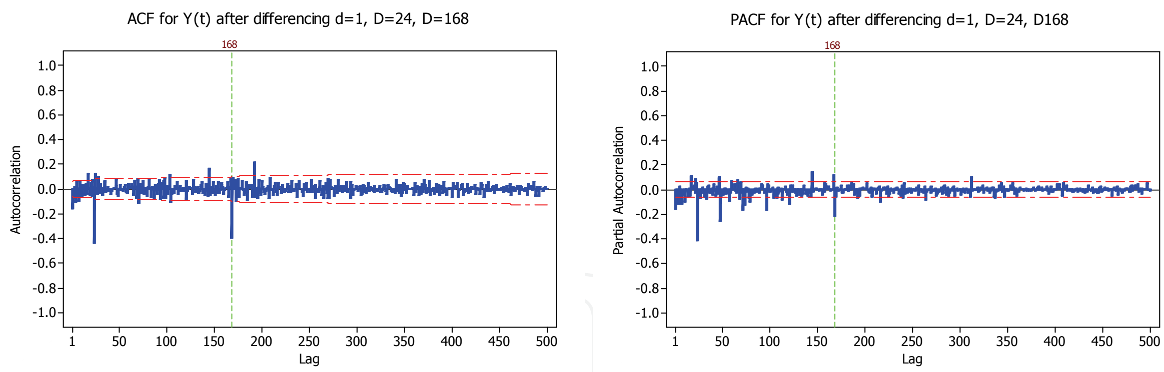


Fig. 6. ACF and PACF for stationary data after differencing d=1, D=24, and D=168.

Then, outlier detection process is only done in the first model, because MSE of this model at in-sample dataset is less than the second model. This process is done iteratively and we find 14 innovational outliers. The first model has out-sample MAPE about 22.8% and the model could be written as

$$(1+0.164B+0.139B^2+0.155B^3+0.088B^4+0.112B^6+0.152B^7+0.077B^9+0.067B^{10}+ \\ 0.069B^{14}+0.089B^{21}+0.072B^{22})(1+0.543B^{168})\ (1-B)(1-B^{24})(1-B^{168})Y_t = \\ (1-0.0674B^8)\ (1-0.803B^{24})a_t\,.$$

Thus, the first model with the outliers is

$$Y_t = \frac{1}{\hat{\pi}(B)}[844I_t^{(830)} - 710.886I_t^{(1062)} + 621.307I_t^{(906)} + -511.067I_t^{(810)} - 485.238I_t^{(1027)} \\ -456.19I_t^{(1038)} + 455.09I_t^{(274)} - 438.882I_t^{(247)} + 376.704I_t^{(1075)} - 375.48I_t^{(971)} + \\ 362.052I_t^{(594)} - 355.701I_t^{(907)} - 329.702I_t^{(623)} + 308.13I_t^{(931)} + a_t]\,,$$

where

$$\hat{\pi}(B) = [(1+0.164B+0.139B^2+0.155B^3+0.088B^4+0.112B^6+0.152B^7+ \\ 0.077B^9+0.067B^{10}+0.069B^{14}+0.089B^{21}+0.072B^{22})(1+0.543B^{168}) \\ (1-B)(1-B^{24})(1-B^{168})]/[(1-0.0674B^8)\ (1-0.803B^{24})]\,.$$

4.2 Result of Elman-Recurrent Neural Network

The Elman-RNN method is applied for obtaining the best network for forecasting electricity consumption in Mengare Gresik. The network elements are the amount of inputs, the amount of hidden units, the amount of outputs, and the activation function in both hidden and output layer. In this research, the number of hidden layers is only one, the activation function in the hidden layer is tangent sigmoid function, and in the output layer is linear function.

The first architecture of Elman-RNN that used for modeling the data is a network with inputs similar to the lags of the best double seasonal ARIMA model. This network uses input lag 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 38, 39, 45, 46, 57, 58, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 182, 183, 189, 190, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 206, 207, 213, 214, 225, 226, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 350, 351, 357, 358, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 374, 375, 381, 382, 393, dan 394. Moreover, the network that was constructed with these input lags is Elman-RNN(101,3,1) and yields MAPE 4.22%.

Then, the second network uses the lags of the best double seasonal ARIMA input and adds 14 detected outliers. These inputs are the lags input as the first network and 14 outliers, i.e. in time period 803th, 1062th, 906th, 810th, 1027th, 1038th, 274th, 247th, 1075th, 971th, 594th, 907th, 623th, and 931th. This network is Elman-RNN(115,3,1) and yields MAPE 4.61%. Furthermore, the third network is network with multiplication of 24 lag input, i.e. inputs are lag 24, 48, …, 480. This third network is Elman-RNN(20,6,1) and yields MAPE 7.55%. Finally, the last network is lag 1 input and multiplication of 24 lag ± 1. The inputs of this fourth network are lag 1, 23, 24, 25, 47, 48, 49, ..., 167, 168, and 169. The network with this inputs is Elman-RNN(22,3,1) and yields MAPE 2.78%.

The forecast accuracy comparison between Elman-RNN models can be seen in Table 2. Based on criteria MSE and MAPE at the out-sample dataset, it can be concluded that Elman-RNN(22,3,1) is the best Elman-RNN for forecasting hourly electricity consumption in Mengare Gresik.

Network	**In-Sample Criteria**			**Out-Sample Criteria**	
	AIC	**SBC**	**MSE**	**MAPE**	**MSE**
RNN(101,3,1)	11.061	12.054	9778.1	4.2167	17937.0
RNN(115,3,1)	10.810	12.073	6755.1	4.6108	21308.0
RNN(20,6,1)	11.468	11.413	22955.0	7.5536	44939.0
RNN(22,3,1)	10.228	9.606	8710.7	2.7833	6943.2

Table 2. The values of each selection criteria of Elman-RNN models

4.3 Comparison between Double Seasonal ARIMA and Elman-RNN

The result of forecast accuracy comparison between double seasonal ARIMA model with and without outliers detection shows that the best model for hourly electricity consumption data forecasting in Mengare is ARIMA([1,2,3, 4,6,7,9,10,14,21,33],1,8)$(0,1,1)^{24}(1,1,0)^{168}$. Then, the comparison is also done with Elman-RNN models. The graphs of the comparison among forecasted values and residuals for the out-sample dataset can be seen in Figure 7. These results show that the residual of Elman-RNN is near to zero compared with ARIMA model. Moreover, the results also show that the forecasted values of Elman-RNN is more accurate than ARIMA model.

In addition, the comparison of forecast accuracy is also done for iterative out-sample MAPE and the result is shown in Fig. 8. This figure shows that Elman-RNN(22,3,1) gives less forecast errors than double seasonal ARIMA and other Elman-RNN models. Hence, all the results of the forecast accuracy comparison show that Elman-RNN yield more accurate forecasted values than double seasonal ARIMA model for electricity consumption data in Mengare Gresik

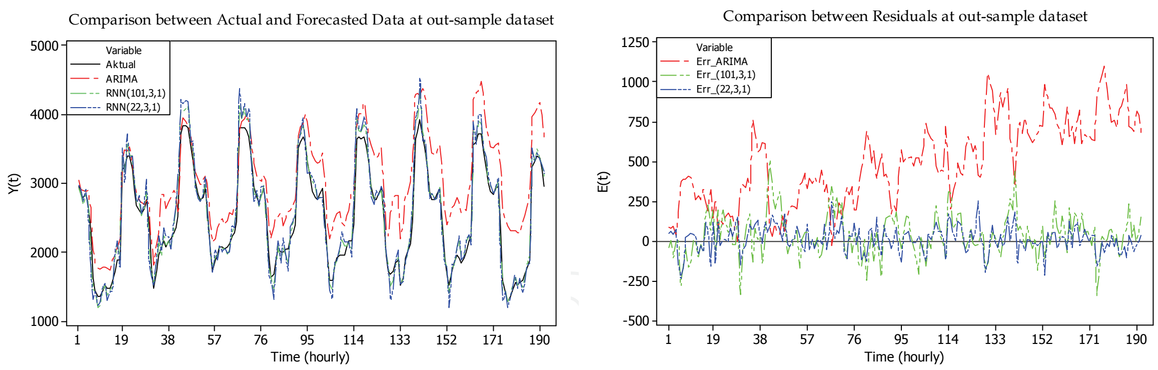


Fig. 7. The comparison of forecast accuracy between ARIMA, Elman- RNN(101,3,1), and Elman-RNN(22,3,1) model.

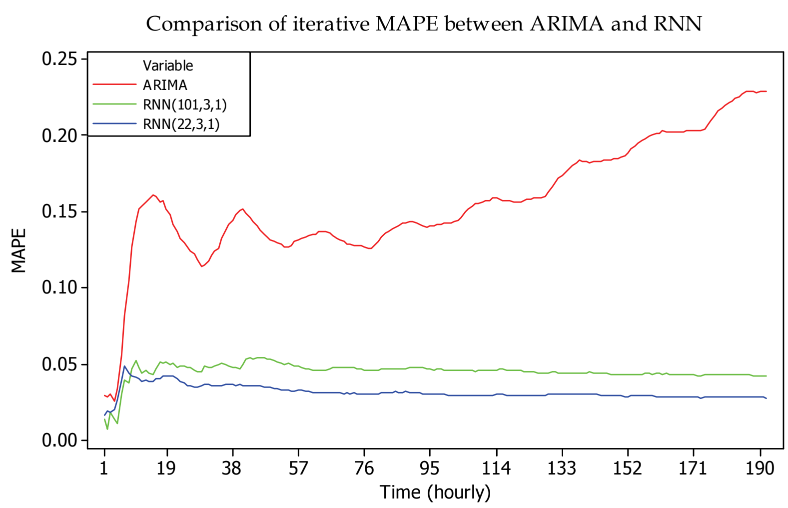


Fig. 8. The comparison of iterative MAPE at out-sample dataset.

5. Conclusion and future work

In this paper, we have discussed the application of RNN for forecasting double seasonal time series. Due to the selection of the best inputs of RNN, the identification of lags input based on double seasonal ARIMA could be used as one of candidate inputs. Moreover, the pattern of the data and the relation to the appropriate lags of the series are important information for determining the best inputs of RNN for forecasting double seasonal time series data. Short-term electricity consumption in Mengare Gresik, Indonesia has been used to compare the forecasting accuracy between RNN and ARIMA models.

The results show that the best order of ARIMA model for forecasting these data is ARIMA ([1-4,6,7,9,10,14,21,33],1,8)$(0,1,1)^{24}(1,1,0)^{168}$ with MSE 11417.426 at in-sample dataset, whereas the MAPE at out-sample dataset is 22.8%. Meanwhile, the best Elman-RNN to forecast hourly short-term electricity consumption in Mengare Gresik is Elman-RNN(22,3,1) with inputs lag 1, 23, 24, 25, 47, 48, 49, 71, 72, 73, 95, 96, 97, 119, 120, 121, 143, 144, 145, 167, 168, and 169, and activation function in the hidden layer is tangent sigmoid function and in the output layer is linear function. This RNN network yields MAPE 3% at out-sample dataset. Hence, the comparison of forecast accuracy shows that Elman-RNN method, i.e. Elman-

RNN(22,3,1), yields the most accurate forecast values for hourly electricity consumption in Mengare Gresik.

In addition, this research also shows that there is a restriction in statistics program, particularly SAS which has facility to do outlier detection. Up to now, SAS program unable to be used for estimating the parameters of double seasonal ARIMA model with adding outlier effect from the outlier detection process. This condition gives opportunity to do a further research related to the improvement of facility at statistics program, especially for double seasonal ARIMA model that involves many lags and the outlier detection.

6. References

Beale, R. & Finlay, J. (1992). *Neural Networks and Pattern Recognition in Human-Computer Interaction*. Ellis Horwood, ISBN:0-136-26995-8, Upper Saddle River, NJ, USA.

Box, G.E.P., Jenkins, G.M. & Reinsel. G.C. (1994). *Time Series Analysis Forecasting and Control, 3rd edition*. Prentice Hall, ISBN: 0-130-60774-6, New Jersey, USA.

Chen, J.F., Wang, W.M. & Huang, C.M. (1995). Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting. *Electric Power Systems Research*, 34, 187-196.

Chong, E.K.P. & Zak, S.H. (1996). *An Introduction to Optimization*. John Wiley & Sons, Inc., ISBN: 0-471-08949-4, New York, USA.

Endharta, A.J. (2009). *Forecasting of Short Term Electricity Consumption by using Elman-Recurrent Neural Network*. Unpublished Final Project, Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia.

Husen, W. (2001). *Forecasting of Maximum Short Term Electricity Usage by implementing Neural Network*. Unpublished Final Project, Department of Physics Engineering, Institut Teknologi Sepuluh Nopember, Indonesia.

Kalaitzakis, K., Stavrakakis, G.S. & Anagnostakis, E.M. (2002). Short-term load forecasting based on artificial neural networks parallel implementation. *Electric Power Systems Research*, 63, 185-196.

Kiartzis S.J., Bakirtzis, A.G. & Petridis, V. (1995). Short-term loading forecasting using neural networks. *Electric Power Systems Research*, 33, 1-6.

Ristiana, Y. 2008. *Autoregressive Neural Network Model (ARNN) for Forecasting Short Term Electricity Consumption at PT. PLN Gresik*. Unpublished Final Project, Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia.

Suhartono. (2007). *Feedforward Neural Networks for Time Series Forecasting*. Unpublished PhD Dissertation, Department of Mathematics, Gadjah Mada University, Indonesia.

Taylor, J.W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54, 799–805.

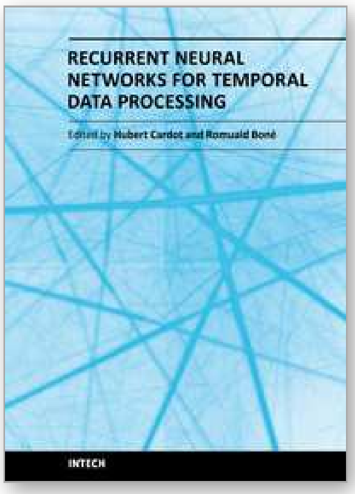
Tamimi, M. & Egbert, R. (2000). Short term electric load forecasting via fuzzy neural collaboration. *Electric Power Systems Research*, 56, 243-248.

Taylor, J.W., Menezes, L.M. & McSharry, P.E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22, 1-16.

Topalli, A.K. & Erkmen, I. (2003). A hybrid learning for neural networks applied to short term load forecasting. *Neurocomputing*, 51, 495-500.

Trapletti, A. (2000). *On Neural Networks as Statistical Time Series Models*. Unpublished PhD Dissertation, Institute for Statistics, Wien University.

Wei, W.W.S. (2006). *Time Series Analysis: Univariate and Multvariate Methods*. 2nd Edition, Addison Wesley, ISBN: 0-321-32216-9, Boston, USA.



Recurrent Neural Networks for Temporal Data Processing

Edited by Prof. Hubert Cardot

ISBN 978-953-307-685-0

Hard cover, 102 pages

Publisher InTech

Published online 09, February, 2011

Published in print edition February, 2011

The RNNs (Recurrent Neural Networks) are a general case of artificial neural networks where the connections are not feed-forward ones only. In RNNs, connections between units form directed cycles, providing an implicit internal memory. Those RNNs are adapted to problems dealing with signals evolving through time. Their internal memory gives them the ability to naturally take time into account. Valuable approximation results have been obtained for dynamical systems.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821